

# Dimension embedding for big data in radio interferometry

S. Vijay Kartik<sup>†</sup>, Rafael E. Carrillo<sup>†</sup> and Yves Wiaux<sup>\*</sup>

<sup>†</sup> Institute of Electrical Engineering, Ecole polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

<sup>\*</sup> Institute of Sensors, Signals & Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

**Abstract**—As radio telescope data is now increasingly high-dimensional, data reduction has become essential to reduce computational load while preserving accurate signal reconstruction. Gridding the continuous Fourier visibilities represents the standard approach to dimension reduction in radio interferometric imaging. This abstract describes a novel dimension embedding technique relying on the multiplication of the data vector by a fat matrix whose entries are drawn from an i.i.d. random Gaussian distribution. Preliminary results suggest that this approach may provide significant improvement in the reduction of data size with respect to standard gridding, for a given target imaging quality.

High-dimensional data acquisition from next-generation radio interferometers can be modelled through the measurement equation  $y = \Phi x + n$ , where  $y \in \mathbb{C}^M$  represents the vector of continuous Fourier measurements (visibilities) corrupted by additive noise  $n \in \mathbb{C}^M$ ,  $\Phi \in \mathbb{C}^{M \times N}$  is the measurement operator, and  $x \in \mathbb{C}^N$  is the underlying signal, with  $M \gg N$ .

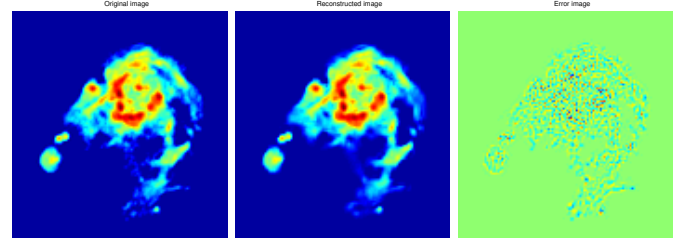
The aim of dimension embedding is reducing data size to minimise computing load, in terms of memory usage and running time of the reconstruction algorithm. Dimension embedding is performed through an embedding matrix  $R \in \mathbb{C}^{M_L \times M}$ ,  $M_L \leq N \ll M$ , leading to an embedded forward model  $y' = \Phi' x + n'$ , with  $y' = Ry$ ,  $n' = Rn$ , and  $\Phi' = R\Phi$ . The gist of dimensional embedding is that the full embedded measurement operator  $\Phi'$  is precomputed once and then stored for further use, which avoids calculations involving the large dimension  $M$  during reconstruction, only dealing with the embedded measurement vector of dimension  $M_L$ . Once the data size is reduced below the image dimension, reconstruction fits into a Compressed Sensing (CS) problem, where the main issues to address are (i) conditions that the measurement operator needs to satisfy, e.g. the Restricted Isometry Property (RIP), and (ii) reconstruction algorithms including adequate prior signal information.

The following dimension embedding techniques are being explored: (i) Holographic embedding [1] involves a ‘gridding down’ of the measurements from the continuous visibilities. In this case,  $R$  identifies with a gridding operator and  $M_L = N$ . Gridding is the standard approach to dimension embedding in radio interferometry; (ii) Random Gaussian embedding involves a matrix  $R$  with zero mean i.i.d. Gaussian random entries, and  $M_L$  can take any arbitrary value, typically  $M_L \leq N$ . For random Gaussian embedding,  $\Phi'$  becomes a Gaussian operator, which might actually approach the characteristics of the optimal sensing matrices promoted by the CS theory [2]. This calls for further theoretical study of the RIP satisfied by  $\Phi'$ , in comparison to  $\Phi$ . Another important advantage of this embedding is that the entries of the embedded noise vector  $n'$  can be shown, on average over realisations of  $R$ , to be an i.i.d. Gaussian distribution, even when the original noise vector  $n$  is not. In this context, the log-likelihood for the data takes the standard form of an  $\ell_2$ -norm term following a  $\chi^2$  distribution with  $M_L$  degrees of freedom. The noise behaviour is therefore very well controlled analytically after embedding, and natural weighting becomes superfluous.

We provide a preliminary comparison between random Gaussian and holographic embeddings on simulated data using the PURIFY toolkit [3], which resorts to convex optimisation for image re-

construction, more specifically to the recently proposed algorithm SARA [4]. From the above considerations, a bound on the  $\ell_2$ -norm data term used by SARA can be set analytically for the random Gaussian embedding, while empirical evaluations are necessary for the holographic embedding.

We use an  $N = 128 \times 128$  model image of the M31 Galaxy (see figure, left plot—shown in log scale), from which  $M = 10N$  continuous visibilities are sampled following a variable density profile with Gaussian shape in the Fourier plane. The input SNR, defined as  $\text{ISNR} = 20 \log_{10}(\|y_0\|_2 / \|n\|_2)$  with  $y_0 = \Phi x$  being the clean measurement vector, is set to 30dB. The reconstruction quality is measured in terms of the output SNR defined as  $\text{OSNR} = 20 \log_{10}(\|x\|_2 / \|x - \hat{x}\|_2)$ ,  $\hat{x}$  being the reconstructed signal. Our simulations show that an OSNR of more than 32dB is reached in the absence of embedding. Holographic embedding achieves an OSNR of around 27dB with, by construction,  $M_L = N$ , while random Gaussian embedding achieves the same imaging quality from an (approximately) three times lower embedding dimension  $M_L = 0.3N$ . The corresponding imaging quality is illustrated in the figure through the reconstructed image (centre, log scale) and the error image (right, linear scale).



If confirmed and extended to truly high-dimensional data, such conclusions may significantly impact the field of radio interferometric imaging by providing significant reduction of memory requirements and computing time. Note that fast implementations of  $\Phi'$  as an operator are critical as it is used at each iterative step during reconstruction. Gaussian embedding—although possibly optimal in terms of imaging quality—remains computationally prohibitive, and in this context, dimension embedding with other random matrices will be further explored, beginning with Bernoulli ensembles (which are expected to be computationally ‘lighter’ as compared to Gaussian ensembles, owing to their sparser and binary structure), and subsampled Hadamard transforms (for which fast implementations exist), as outlined in [5].

## REFERENCES

- [1] I. S. Sullivan et. al., “Fast holographic deconvolution: A new technique for precision radio interferometry,” *The Astrophysical Journal*, 2012.
- [2] H. Rauhut, “Compressive sensing and structured random matrices,” *Radon Series Comp. Appl. Math.*, 2011.
- [3] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, “PURIFY: a new approach to radio-interferometric imaging,” *MNRAS*, 2014.
- [4] —, “Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging,” *MNRAS*, 2012.
- [5] J. A. Tropp, “Improved analysis of the subsampled randomized hadamard transform,” *Adv. Adapt. Data Anal.*, 2014.